Mol Gen Genomics (2006) 275: 479–491
DOI 10.1007/s00438-006-0106-z

ORIGINAL PAPER

**James E. Frelichowski Jr · Michael B. Palmer
Dorrie Main · Jeffrey P. Tomkins · Roy G. Cantrell
David M. Stelly · John Yu · Russell J. Kohel
Mauricio Ulloa**

Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends

Received: 9 December 2005 / Accepted: 14 January 2006 / Published online: 25 February 2006

Abstract Fine mapping and positional cloning will eventually improve with the anchoring of additional markers derived from genomic clones such as BACs. From 2,603 new BAC-end genomic sequences from *Gossypium hirsutum* Acala 'Maxxa', 1,316 PCR primer pairs (designated as MUSB) were designed to flank microsatellite or simple sequence repeat motif sequences. Most (1164 or 88%) MUSB primer pairs successfully amplified DNA from three species of cotton with an average of three amplicons per marker and 365 markers (21%) were polymorphic between *G. hirsutum* and *G. barbadense*. An interspecific RIL population developed from the above two entries was used to map 433 marker loci and 46 linkage groups with a genetic distance of 2,126.3 cM covering approximately 45% of the cotton genome and an average distance between two loci of 4.9 cM. Based on genome-specific chromosomes identified in *G. hirsutum* tetraploid (A and D), 56.9% of the coverage was located on the A subgenome while 39.7% was assigned to the D subgenome in the genetic map, suggesting that the A subgenome may be more polymorphic and recombinationally active than originally thought. The linkage groups were assigned to 23 of the 26 chromosomes. This is the first genetic map in which the linkage groups A01 and A02/D03 have been assigned to specific chromosomes. In addition the MUSB-derived markers from BAC-end sequences markers allows fine genetic and QTL mapping of important traits and for the first time provides reconciliation of the genetic and physical maps. Limited QTL analyses suggested that loci on chromosomes 2, 3, 12, 15 and 18 may affect variation in fiber quality traits. The original BAC clones containing the newly mapped MUSB that tag the QTLs provide critical DNA regions for the discovery of gene sequences involved in biological processes such as fiber development and pest resistance in cotton.

Keywords Simple sequence repeat (SSR) · Bacterial artificial chromosome (BAC) · BAC-end sequencing · Genetic mapping · Quantitative trait loci (QTL) · Fine mapping · Molecular tagging

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00438-006-0106-z and is accessible for authorized users.

Communicated by S. Hohmann

J. E. Frelichowski Jr · M. Ulloa
USDA-ARS, W.I.C.S. Res. Unit,
Cotton Enhancement Program, Shafter, CA, USA

M. B. Palmer · D. Main · J. P. Tomkins
Department of Genetics and Biochemistry,
Clemson University, Clemson, SC, USA

R. G. Cantrell
Cotton Incorporated, Cary, NC, USA

D. M. Stelly
Texas A&M University, College Station, TX, USA

J. Yu · R. J. Kohel
USDA-ARS, College Station, TX, USA

M. Ulloa (✉)
17053 N. Shafter Ave, Shafter, CA 93263, USA
E-mail: mulloa@pw.ars.usda.gov
Tel.: +1-661-7468009
Fax: +1-661-7461619

Introduction

Understanding the genetic basis of genome structure and of biological processes for favorable traits is relevant to a range of research objectives and also has applications for improvement of crops. Cotton is a model crop for cytogenetic, genomic and evolutionary biology research. It is also an economically important renewable source of natural fiber and secondary products such as oil, livestock feed and cellulose (Anderson 1999; Gregory et al. 1999). Cotton fibers are thought to be the longest unicellular structures in the plant kingdom and are a model for research on cell elongation and cell wall biosynthesis (Hof and Saha 1997; Kim and Triplett 2001). *Gossypium*,

the cotton genus, contains 45 diploid and 5 allotetraploid species (Fryxell 1992; Percival et al. 1999; Ulloa et al. 2006) grouped into nine genomic types ($x = 2n = 26$, or $n = 13$) with the designations AD, A, B, C, D, E, F, G, and K (Percival et al. 1999). The complexity of the *Gossypium* genomes has afforded research opportunities on the evolution among diploid and allotetraploid species (Brubaker et al. 1999; Jiang et al. 1998), genetic diversity and changes related to domestication (Paterson 2002).

Many traits of interest are likely the product of complex biological processes in which multiple genes and environments interact. Production of cDNA (Arpat et al. 2004) and genomic libraries (Gonzalez et al. 2005; Rong et al. 2004; Tomkins et al. 2001) are important for studying sequence composition of the genome and candidate gene sequences involved in biosynthetic processes. Another strategy for gene discovery is the construction of genetic linkage maps from molecular markers derived from these sources and from other molecular techniques (Ulloa and Meredith 2000; Ulloa et al. 2002; Mei et al. 2004; Nievergelt et al. 2004; Rong et al. 2004; Lacape and Nguyen 2005; Lin et al. 2005; Shen et al. 2005; Ulloa et al. 2005; Wang et al. 2005). Markers tightly linked to genes of interest are valuable for cloning sequences such as retroelements (Taliercio and Ulloa 2003) which impact genome structure and evolution, and loci that affect plant–pathogen interactions (De Kock et al. 2005). For more complex traits genetic maps have been used to partition traits into quantitative trait loci (QTLs) (Lacape et al. 2003; Ulloa et al. 2005). Precise marker location, order and duplication of loci are obstacles to map-based cloning of sequences of interest (Ulloa et al. 2005). Genome sequencing, physical alignment of genomic sequences into the chromosomal maps and anchoring of the genetic maps are all steps that will improve the accuracy of the genetic maps and enable discovery of genes underlying QTLs and their functions in important biological processes in cotton.

To understand the complexities of genetic linkage mapping in cotton, we must first understand the evolution and structure of the cotton genome. Cotton genomics is in its infancy and despite ongoing efforts by the cotton community, a high-density molecular PCR-based map to facilitate gene discovery and marker-assisted selection in the public domain is still lacking. Knowledge of cotton genomic structure remains at least 10 years behind the well-coordinated global efforts in Arabidopsis, rice, maize, *Medicago* and other model systems.

BAC libraries are stable and efficient for cloning genomic DNA and are commonly used to initiate genome sequencing (Shizuya et al. 1992). They are especially useful in that they can provide direct integration of genetic and physical maps (Kim et al. 2005). BACs containing genetic markers can be used to probe intact chromosomes to provide cytogenetic evidence of marker location and order (Islam-Faridi et al. 2002). BAC sequences have also been used in saturating genetic maps (Kim et al. 2005), map-based cloning of genes (Kilian et al. 1997), comparative mapping, and genome analysis (Mao et al. 2000). They also provide another valuable resource for identifying microsatellite markers (Hong et al. 2004).

Microsatellites, alternatively called simple sequence repeats (SSR), have become markers of choice because they are locus specific, often co-dominant, multi-allelic and transferable across species and even genera (Gupta et al. 1996). Microsatellites are arrays of nucleotides from 1 to 6 bases in length and are common components of genes and whole genomes (Tautz and Schotterer 1994). SSR markers have been developed for a range of applications in scientific research including measuring genetic diversity and genetic mapping (Varshney et al. 2005). ESTs derived from developing fibers (Arpat et al. 2004) contain numerous SSRs (Saha et al. 2003; Han et al. 2004; Qureshi et al. 2004; Park et al. 2005) and are useful to assess location and effect of candidate genes in trait expression.

A newer approach to gene discovery in cotton is proposed herein by anchoring existing cotton BAC clones into a chromosomal framework of cotton by mapping and hypoaneuploid assays of discovered SSR markers. A BAC library was assembled from *G. hirsutum* Acala 'Maxxa' with an average insert size of 137 kb ranging from 80 to 275 kb (Tomkins et al. 2001). New sequence information was generated from BAC contigs for this study, and data mining for SSRs were done. One of the objectives of this research is to develop a portable set of polymorphic SSR markers from these Maxxa BAC-end sequences. Herein, the newly developed markers are anchored into the cotton genetic map with publicly available SSR markers and subgenome-specific chromosomes with hypoaneuploid stocks. A limited search for QTLs associated with fiber quality traits with these markers also is performed to assess the ability of the new markers to detect QTLs. BAC-end markers that are associated with QTLs for complex traits will be essential for fine mapping and map-based sequencing of genes to reveal their biological functions. Localization of SSR markers to all 26 chromosomes, and verification of their order and colinearity within their subgenome-specific chromosomes are needed to complete the development of a highly saturated consensus map in cotton. Physical mapping of genetic markers also enables comparisons among *Gossypium* genomes and the detection of deletions, duplications, or rearrangements that may be specific to species, or crosses within or between species.

Materials and methods

Development of BAC-end SSR markers

The BAC library of Maxxa was developed by Tomkins et al. (2001). The end sequences of the BAC clones (Palmer et al. 2004) were screened for the presence of di,

tri, tetra, or pentanucleotide repeat motifs with a minimum length of ten bases. SSRs were selected for non-repetitive sequences of 30–70% GC content flanking the SSRs. From approximately 45,000 sequences, 2,603 were identified as suitable for use in primer design. The first 1,373 sequences were subjected to PCR primer design using the default parameters of Primer3 software (Rozen and Skaletsky 2000) to flank the SSRs and amplify fragment sizes between 125 and 350 base pairs (bp). The parameters, such as maximum fragment size and GC clamp were relaxed to allow for design of primers on more SSR sequences. Eight hundred primer pairs were designed for the first set of SSRs. The remaining 1,230 SSRs were designed with FastPCR software (Kalendar 2005) to flank the SSRs. This software implements updated methods for primer design, and integrates easily with spreadsheet programs to handle large numbers of sequences easily. Five hundred and sixteen SSRs were selected for designing the second set of primer pairs. Each marker from a primer pair was identified with the alpha code MUSB: M for microsatellite, U for the last name of the principal investigator, S for simple sequence repeat and B for BAC. Each primer pair was given a marker number starting with MUSB followed by a number from 1 to 1,316. All primer pairs were synthesized by Proligo LLC (Boulder, CO) and were received desalted. Each primer was dissolved in Tris-base (10 mM) EDTA (0.5 mM) buffer at a concentration of 10 μM. Each PCR reaction consisted of the following: 20 ng of DNA, 1× reaction buffer II (Applied Biosystems (ABI), Foster City, CA), 3 mM $MgCl_2$ (ABI), 0.2 mM each of four dNTPs (dATP, dTTP, dCTP and dGTP, Promega Corporation, Madison, WI), 0.1 μM of each a forward and reverse oligonucleotide primer for flanking an SSR sequence, and 0.5 U of Amplitaq Gold Taq Polymerase (ABI) in a total volume of 15 μl. The reaction mixes of the 800 MUSB (Primer3 designed) primer pairs were subjected to the following PCR protocol in a MJ PTC-200 thermocycler (MJ Research Incorporated, Watertown, MA): an initial activation phase of 95°C for 10 min, a touchdown amplification phase of ten cycles of three consecutive temperatures, denaturation at 94.0°C for 15 s, annealing at 60.0°C for 30 s, and extension at 72.0°C for 90 s, (with the annealing temperature decreased by 0.5°C each cycle), a constant amplification phase of 33 cycles of three temperatures, 94.0°C for 15 s, 55.0°C for 30 s, and 72.0°C for 1 min, a final extension step of 72.0°C for 6 min, and storage at 4°C. For the 516 MUSB (FastPCR designed) primer pairs, the PCR conditions were varied at the touchdown phase with 21 cycles of three temperatures, 94.0°C for 15 s, 60.0°C for 30 s, 72.0°C for 90 s, (with the annealing temperature decreased by 0.5°C each cycle), and the constant amplification phase with 33 cycles of three temperatures, 94.0°C for 15 s, 50.0°C for 30 s.

PCR products were separated on 3% Super Fine Resolution (SFR^{tm}) agarose gels (Amresco, Solon, OH) in 1× TBE buffer at 4 V/cm. Gels were stained with ethidium bromide and visualized with UV illumination and image analysis by Alpha Imager (Alpha Innotech Corp. San Leandro, CA) software v.5.5. Easily viewed, expected and unexpected fragment sizes were estimated in base pairs with a size marker (pBR322 DNA-*Msp I* digest, 3032 L, New England Biolabs Incorporated, Beverly, MA) and scored as alleles.

Plant material

PCR amplification with SSR primers and evaluation of polymorphisms was performed on template DNA of four cotton genotypes, *G. hirsutum* Acala Maxxa, *G. hirsutum* cv.Texas Marker-1 (TM-1), *G. barbadense* acc. 3-79 and *G. raimondii* ($D_{5\text{-}3}$) (Table 1). Maxxa was used as a control for PCR amplification and to check the amplicons for expected sizes, TM-1 and 3-79 were used to find polymorphic markers for further use in genetic mapping and $D_{5\text{-}3}$ was used to extend polymorphism to other cotton species. Markers polymorphic between *G. hirsutum* and *G. barbadense* 3-79 were assayed on a panel of 16 entries that included four different tetraploid species (Table 1) to calculate PIC values using the method of Anderson et al. (1993) and further described by Park et al. (2005). PIC values are a method of comparing molecular markers based on the number of alleles and their combinations in a population (Botstein et al. 1980).

Table 1 List of cotton cultivars and accessions used to study MUSB microsatellite marker variation

Cultivar/accession[a]	Taxonomy	Genome
DNA sample panel 1		
Acala Maxxa	*G. hirsutum* L.	AD_1
TM-1	*G. hirsutum* L.	AD_1
Pima 3-79	*G. barbadense* L.	AD_2
($D_{5\text{-}3}$)	*G. raimondii* L.	$D_{5\text{-}3}$
DNA sample panel 2		
MD51ne-okra	*G. hirsutum* L.	AD_1
MD51ne-nomal	*G. hirsutum* L.	AD_1
HQ95-6	*G. hirsutum* L.	AD_1
Acala Maxxa	*G. hirsutum* L.	AD_1
Phytogen 72	*G. hirsutum* L.	AD_1
Shafter-SC01	*G. hirsutum* L.	AD_1
SL-1-7-1	*G. hirsutum* L.	AD_1
9456-0 Naked seed	*G. hirsutum* L.	AD_1
Ligon lintless	*G. hirsutum* L.	AD_1
TM-1	*G. hirsutum* L.	AD_1
NM24016	Interspecific inbred (*Gh* × *Gb*)	
Pima S-7	*G. barbadense* L.	AD_2
03MUGH3	*G. barbadense* L.	AD_2
Pima 3-79	*G. barbadense* L.	AD_2
WT936	*G. tomentosum* Nuttall ex Seemann	AD_3
W400	*G. mustelinum* Miers ex Watt	AD_4

[a]DNA sample panel 1 was used to test PCR amplification and prescreen polymorphisms for 1316 MUSB markers. The primer-pairs revealing polymorphism between *G. hirsutum* L. and *G. barbadense* L. in the DNA sample panel 1 were tested further using DNA sample panel 2. The PIC values were calculated using the information of allelic variation from the second DNA sample panel

Chromosomal assignment of SSR markers

Chromosomal assignment of markers was accomplished with hypoaneuploid F_1 stocks developed from an interspecific cross between TM-1 and 3-79 (Stelly 1993). These stocks are maintained at the Cotton Cytogenetics Collection at Texas A&M University, and are frequently used for chromosomal assignment of select SSR markers (Stelly 1993; Liu et al. 2000; Ulloa et al. 2005). The cytogenetic stocks consisted of 15 primary monosomic and 25 monotelodisomic lines where an entire chromosome, all or most of a chromosome arm, or several different specific chromosome segments of TM-1 are missing. The intact homologous chromosomes of 3-79 are all present. A set of primer pairs and their marker loci that were easily scored on the RILs and generated a TM-1 amplicon that was different in size from 3-79 were assayed on the DNA of the hypoaneuploid F_1 stocks. Microsatellite markers were assigned to the deficient chromosome of the aneuploid stock from which the TM-1 allele was not amplified. Microsatellite loci localized to one of the chromosomes 1 to 13 were assigned to the A-subgenome (At), whereas loci localized to chromosomes 14 to 26 were assigned to the D-subgenome (Dt).

Genetic map construction

The markers with the best resolution under the agarose system were selected. A set of 183 recombinant inbred lines (RILs) from an interspecific cross between TM-1 and 3-79 was used as template for creation of genetic linkage groups for these SSR marker loci (Park et al. 2005). Two hundred twenty-one MUSB markers polymorphic between TM-1 and 3-79 were selected for mapping. SSR marker data (67 MUSS, 42 MUCS and 70 BNL) used in the creation of the map of Park et al. (2005), and additional BNL (25) JESPR (9) NAU (22) and CIR (2) primer pairs assayed on the RILs, were combined with the MUSB marker data. Each RIL was scored for the presence or absence of the TM-1 and 3-79 alleles for each marker locus with each primer pair. For each locus alleles were considered codominant unless the frequency of heterozygotes were high or multiple parental alleles did not co-segregate, in which cases each marker allele was scored as a dominant marker locus. Multiple or duplicated loci detected by a single primer-pair were designated by adding a lowercase letter in alphabetical order after the name of the primer-pair with the letter 'a' being the largest amplicon, the letter 'b' the next largest, etc. The data were coded for analysis and linkage grouping with JoinMap® 3.0 (Stam and Van Ooijen 1995). Codominant amplicons were coded as A, B or H for presence of TM-1 amplicon only or the 3-79 amplicon only (i.e. homozygous), or for the presence of both (i.e. heterozygous), respectively. Dominant amplicons of TM-1 were coded as A for the presence of the TM-1 amplicon or B for absence, and likewise the dominant amplicons of 3-79 were coded as B for presence or A as absence. Missing data in both cases were coded as U. The Kosambi map function (Kosambi 1944) was used to convert recombination frequency to genetic map distance (centiMorgan, cM). Most of the linkage groups were determined at LOD scores ≥6. Deviation from a 1:1 segregation ratio expected for RILs was tested by Chi-square statistics for each marker.

Gene function annotation of SSR containing BAC-end sequences

The original BAC-end sequences of 131 MUSB markers were searched for protein homologies in Genebank using a BLASTX search (Altschul et al. 1990). The best matching sequence ($P < 0.001$) was obtained to assign putative gene function to the cotton genomic sequence.

QTLs analysis of fiber quality

The fiber quality data of the F_2 progenitors (Kohel et al. 2001) of the RILs used in this study consisted of fiber elongation (*Ef*), fiber bundle strength (*Sf*), fiber length at 2.5% (2.5% *Lf*) and 50% (50% *Lf*) and fiber fineness (*Ff*). The phenotypic data from the RILs were not available at the time of this study but with F_2 data and an SSR marker map of Park et al. (2005), expanded with BAC-end derived SSR markers, we sought to compare the chromosomal locations of previously reported QTLs (Kohel et al. 2001; Park et al. 2005) by using RIL DNA and SSR markers. QTL analysis was conducted using MapQTL 4.0 (Van Ooijen and Maliepaard 1996) with interval mapping, MQM and restricted MQM mapping procedures. Threshold values for LR were determined empirically after 1,000 permutation test for all traits (Churchill and Doerge 1994).

Results

Microsatellite identification and marker design

A set of 2,603 SSR sequences were identified and 1,316 were suitable for design of flanking PCR primers. The 2,603 are only a select subset of the BAC-end sequencing output from approximately 45,000 BAC-end sequences and data mining done for SSRs with a minimum length of ten bases; therefore, data and inferences about SSR composition and frequency are limited to those contained in the MUSB markers. The relative abundance of the four repeat types among the 1,316 SSRs were: 489 di, 263 tri, 490 tetra and 74 pentanucleotides. The most common dinucleotide motifs were AT and TA, while TCT, AAG and AAT were most common for trimeric SSRs. For tetra- and pentanucleotide repeat motifs, TTTA, TTAA, TTTC and AAAAT, TTTTC, and TTATT, respectively, were observed at the highest frequencies, respectively. The most consistent observation

was the rare occurrence of the nucleotides C and G together in the same motif. Only 1164 MUSB markers could be scored for PCR amplification products, and Table 2 shows the numbers of MUSB markers that produced all monomorphic PCR fragments or at least one polymorphic allele for flanking di, tri, tetra, and pentanucleotide SSRs and for each SSR size in number of repeat units. The minimum number of repeat units to qualify for SSR marker design was 5, 4, 3, and 3 for di, tri, tetra and pentanucleotides, respectively. Variability in the number of repeat units resulted in greater proportions of polymorphic markers. All of the motifs generally followed these patterns (data not shown), but in particular the two motifs AATT and AAAAT with 9 of 10 and 6 of 6 markers, respectively, had high numbers of polymorphic bands(s).

MUSB marker polymorphisms

Only 1,316 SSR sequences were suitable for design of flanking PCR primer pairs due to insufficient length and poor quality of flanking sequences that resulted in primers annealing to one another or falling outside the desirable TM range. The amplicons of a primer pair applied to the species in Panel 1 (Table 1) were categorized according to success or failure of the reaction to produce visible and easily scored alleles or the presence of only non-specific amplification (NSA) that did not correspond to the expected size fragment from the original BAC-end sequence and were not resolved in our agarose electrophoresis system. Each visible and consistent amplicon of the markers was treated as a marker allele and polymorphisms were recorded according to the presence of a different size allele or the absence of an allele in one or more genotypes. MUSB primer pairs that failed these criteria were 146 (11.1% of the total), while six (0.5%) produced only non-specific reaction products and precluded the differentiation of alleles. Useful amplification was achieved with the remaining 1,164 MUSB markers (88.4%), and they were classified further into those that produced only monomorphic PCR products (491, 37.3%) and those that produced at least one polymorphic fragment (673, 51.2%). Within both marker categories the proportion of MUSB markers that produced alleles of the expected or only unexpected sizes, were 481 and 10, respectively, for monomorphic markers and 524 and 149, respectively, for polymorphic markers. Many reactions that produced alleles of the expected sizes also amplified additional alleles of unexpected sizes and NSA. Table 3 summarizes the number of markers producing at least one polymorphic or only all monomorphic alleles among the four genotypes using 663 Primer3 and 501 FastPCR designed MUSB primers. More than half of the markers were polymorphic between the tetraploid and diploid species ($D_{5\text{-}3}$). Three hundred seventy-one (31.9%) markers were polymorphic between the tetraploid species, but only 132 markers differentiated the two *G. hirsutum* genotypes. The estimated number of amplicons (and average per marker) for each genotype was 1,928 (2.9), 2,015 (3.0), 2,000 (3.0) and 1,924 (2.9) for Maxxa, TM-1, 3-79 and $D_{5\text{-}3}$, respectively. The amplification of the MUSB markers across just these three species suggests high transferability of the markers in cotton germplasm.

In a comparison of the two primer design methods with all types of alleles, 541 (67.6%) and 105 (13.1%) Primer3 designed markers produced expected and NSA alleles, respectively, whereas 464 (89.9%) and 27 (5.2%) FastPCR designed markers produced expected and NSA alleles, respectively. Fewer failed reactions were observed for FastPCR designed primers (13) than for Primer3 designed primers (133).

PIC values were determined for MUSB markers polymorphic between TM-1 and 3-79 based on 16 DNA samples in order to assess molecular diversity and potential utility of these markers for fingerprinting and genetic mapping. PIC values ranged from 0.12 to 0.85, with an average of 0.4. MUSB0580 showed the highest PIC value, while the lowest value was found in three markers. The markers with higher PIC values possess greater potential to reveal allelic variations and, therefore, can be preferentially used to aid in choosing appropriate parental lines for molecular mapping or assessing genetic diversity within a population (Anderson et al. 1993).

Table 2 Number of monomorphic or polymorphic MUSB markers classified by their SSR repeat motif composition as di-, tri-, tetra-, or pentanucleotides and number of repeat units

Added units	Di		Tri		Tetra		Penta		Sum	
	Mono	Poly	Mono	Poly	Mono	Poly	Mono	Poly	Mono	Poly
Min.[a]	124	155	85	95	172	231	25	37	406	517
≥1	38	48	13	19	11	12	1	4	63	83
≥2	12	17	3	5	2	5	0	3	17	30
≥3	3	11	0	2	1	2	0	1	4	16
≥4	1	14	0	9	0	3	0	1	1	27
Totals	178	245	101	129	186	253	26	46	491	673
	423		230		439		72		1164	

[a]Minimum number of repeat units per motif size: di = 5, tri = 4, tetra = 3, penta = 3, each row represents the addition of one more unit to the minimum SSR size

Table 3 1164 MUSB markers classified according to polymorphic or monomorphic pairwise comparisons among four cotton genotypes

		Maxxa	Percent	TM-1	Percent	3-79	Percent
Poly	TM-1	132	11.3				
	3-79	371	31.9	365	31.4		
	$D_{5\text{-}3}$	610	52.4	610	52.4	618	53.1
Mono	TM-1	1032	88.7				
	3-79	793	68.1	799	68.6		
	$D_{5\text{-}3}$	554	47.6	554	47.6	546	46.9

Percent based on denominator of 1,164 (total markers)
Maxxa = *G. hirsutum* Acala 'Maxxa'; TM-1 = *G. hirsutum* TM-1; 3-79 = *G. barbadense* 3-79; $D_{3\text{-}5}$ = *G. raimondii* ($D_{5\text{-}3}$)

Chromosome location for microsatellite DNA fragments

Primers for a subset of 35 MUSB and 20 unassigned BNL markers were used to amplify DNA fragments from 46 F_1 substitution aneuploid lines to assign the chromosomal location of the markers (Liu et al. 2000; Stelly 1993; Ulloa et al. 2005). The unambiguous absence of TM-1 fragments from the aneuploid stocks was observed for 20 microsatellite markers (Figs. 1 and 2). These markers were assigned to 17 chromosomes (Chr.1, 2, 4, 5, 8, 9, 10, 11, 12, 15, 16, 17, 18, 19, 20, 22 and 23) and of these, 11 could be assigned to a chromosome arm (Chr.1*sh*, 8*sh*, 10*sh*, 10*Lo*, 11*sh*, 15*sh*, 16*Lo*, 17*Lo*, 20*sh*, 20*Lo* and 22*sh*). The chromosomal assignment of the new MUSB and BNL markers were helpful in confirming the chromosomal assignment of their linkage groups with prior mapping information. In addition, the chromosomes 8, 19 and 22 are being reported here as the linkage groups with MUSS and MUCS markers that were not assigned in Park et al. (2005). A linkage group was assigned to chromosome 8 on the basis of five markers (MUSB0100, MUSS250b, MUSB0818a, MUCS148, and MUSB0442) that were absent for the hypoaneuploid line corresponding to a missing short arm of chromosome 8. Two linkage groups are assigned to chromosome 19 solely on the basis of the aneuploid assays. Chromosome 11 is assigned to a linkage group based on the hypoaneuploid assays with MUCS557 and MUSS123 along with the assignment of JESPR135 to this chromosome (Park et al. 2005).

Genetic mapping

Two hundred twenty-one MUSB markers were genotyped on the RILs but only 131 were selected for mapping due to ease of allele scoring. The chromosomal location of SSR markers were derived from prior mapping or the hypoaneuploid assays and assisted in the creation of linkage groups with LOD scores equal to or greater than 6. The genetic linkage map (Figs. 1 and 2) consisted of 433 loci from the new 125 MUSB (169), 67 MUSS (74), 42 MUCS (48), 95 BNL (101), 22 NAU, 9 JESPR, and 8 CIR (10) primer sets. Primer sequences of the new MUSB markers are listed in Table S1. The map covers a total of 2126.3 centiMorgan (cM) units covering approximately 45% of the cotton genome (Reinisch et al. 1994; Stelly 1993). The map is composed of 46 linkage groups, with 39 representing portions of 23 chromosomes and 7 as A03, D08 or 5 unassigned linkage groups with the unassigned linkage groups having a combined genetic distance of 70.4 cM (Fig. 2). The average distance between marker loci is 4.9 cM with 36.1 cM as the largest distance and 0.14 cM as the smallest distance. The At subgenome (Fig. 1) consists of 249 total marker loci (including 107 new MUSB marker loci) with a genetic distance of 1,210.8 cM (average of 4.9 cM between loci), and the Dt subgenome (Fig. 2) consists of 168 marker loci (including 59 MUSB marker loci) with a genetic distance of 845.1 cM. (average of 5.0 cM between loci).

Segregation distortion was evident for only 70 (17%) markers (using a Chi-square value threshold of 10.83, $P < 0.001$). At $P < 0.001$ and $P < 0.05$ the proportion of marker loci skewed towards TM-1 allele is (66.7%) and (72.7%), respectively. Twenty-four of these distorted markers mapped to the same chromosome as with the hypoaneuploid assays and is a justification of inclusion of these markers in the map. For example on chromosome 7 a cluster of six markers shows high segregation distortion but five of them are assigned by means of hypoaneuploid deficiency analysis to chromosome 7 and another CIR141 also maps to chromosome 7 in another study (Nguyen et al. 2004). An additional 18 markers had distorted segregation but ten of them were assigned to chromosome 15 by means of hypoaneuploid deficiency analysis, and all five BNL markers in this cluster agree with locations in another published map (Nguyen et al. 2004). High segregation distortion of marker loci also occurred in an intraspecific map within *G. hirsutum* (Shen et al. 2005).

The chromosomal assignments to our linkage groups generally agree with other published cotton SSR maps (Nguyen et al. 2004; Rong et al. 2004). Several modifications are proposed that merge linkage groups and assign markers to chromosomes with hypoaneuploid deficiency analysis. Markers normally reported as linkage group A02 are linked with markers MUSB0100, MUSB0442, MUSS250b and MUCS148 that were associated with the short arm of chromosome 8 according to hypoaneuploid deficiency analysis. Therefore, we assign this linkage group to chromosome 8. Several markers associated with linkage group A02 in

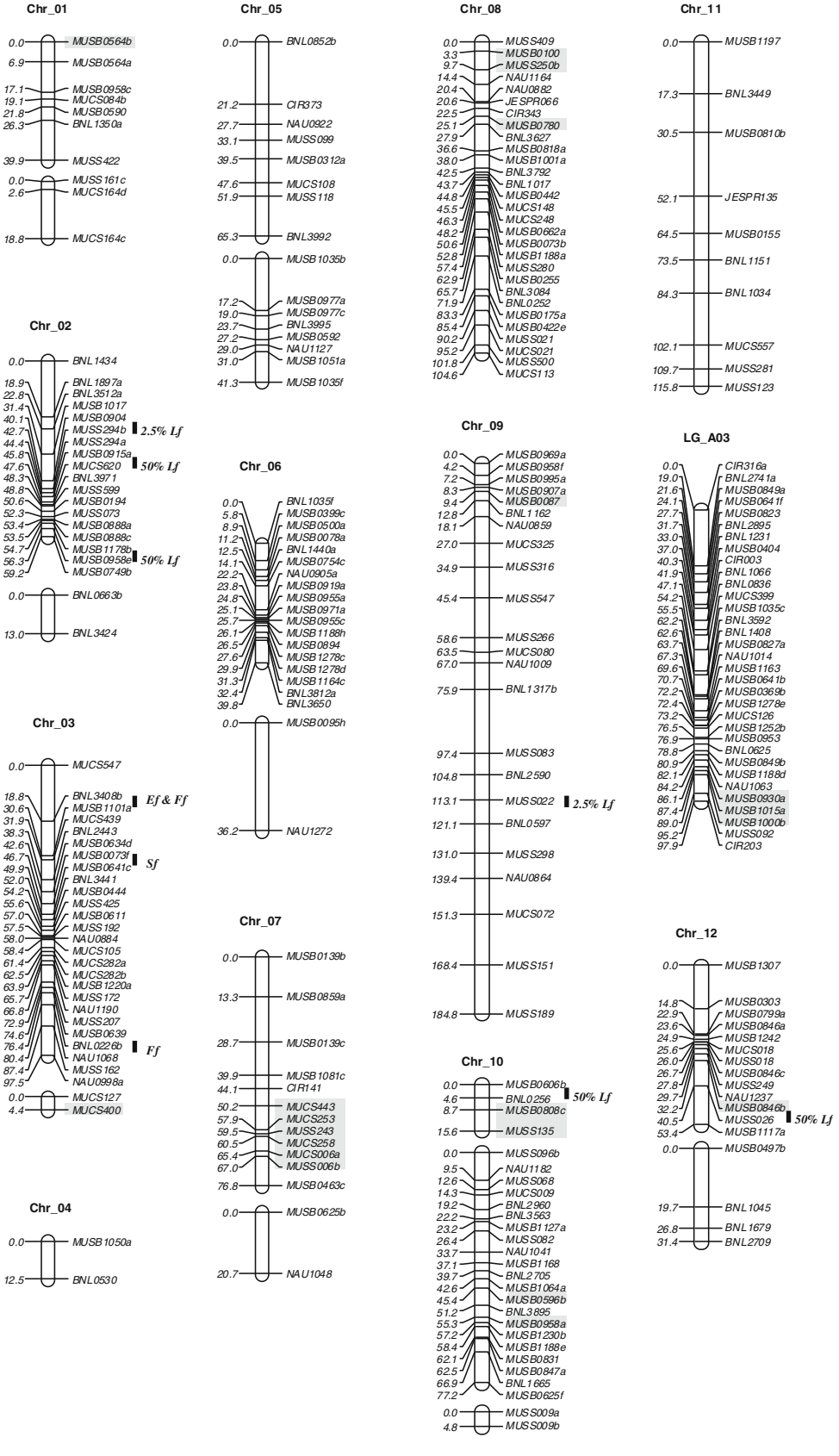

Fig. 1 A subgenome coverage from different chromosomes; the cotton genetic linkage map was constructed from 183 recombinant inbred lines (RILs) from the interspecific cross *Gh cv.* TM1 × *Gb acc.* Pima 3-79. from a total of 433 loci and 46 linkage groups, including 169 new BAC-MUSB loci and framework SSR markers (BNL, JESPR, CIR and NAU) to anchor the map. The markers with significant segregation distortion (Chi-square > 10) are indicated by *gray blocks*. Map intervals for putative QTL for five fiber quality-related traits, fiber elongation (*Ef*), fiber bundle strength (*Sf*), fiber length at 2.5% (*2.5% Lf*), fiber length at 50% (*50% Lf*), and fiber fineness (*Ff*), were presented by *solid vertical bars*

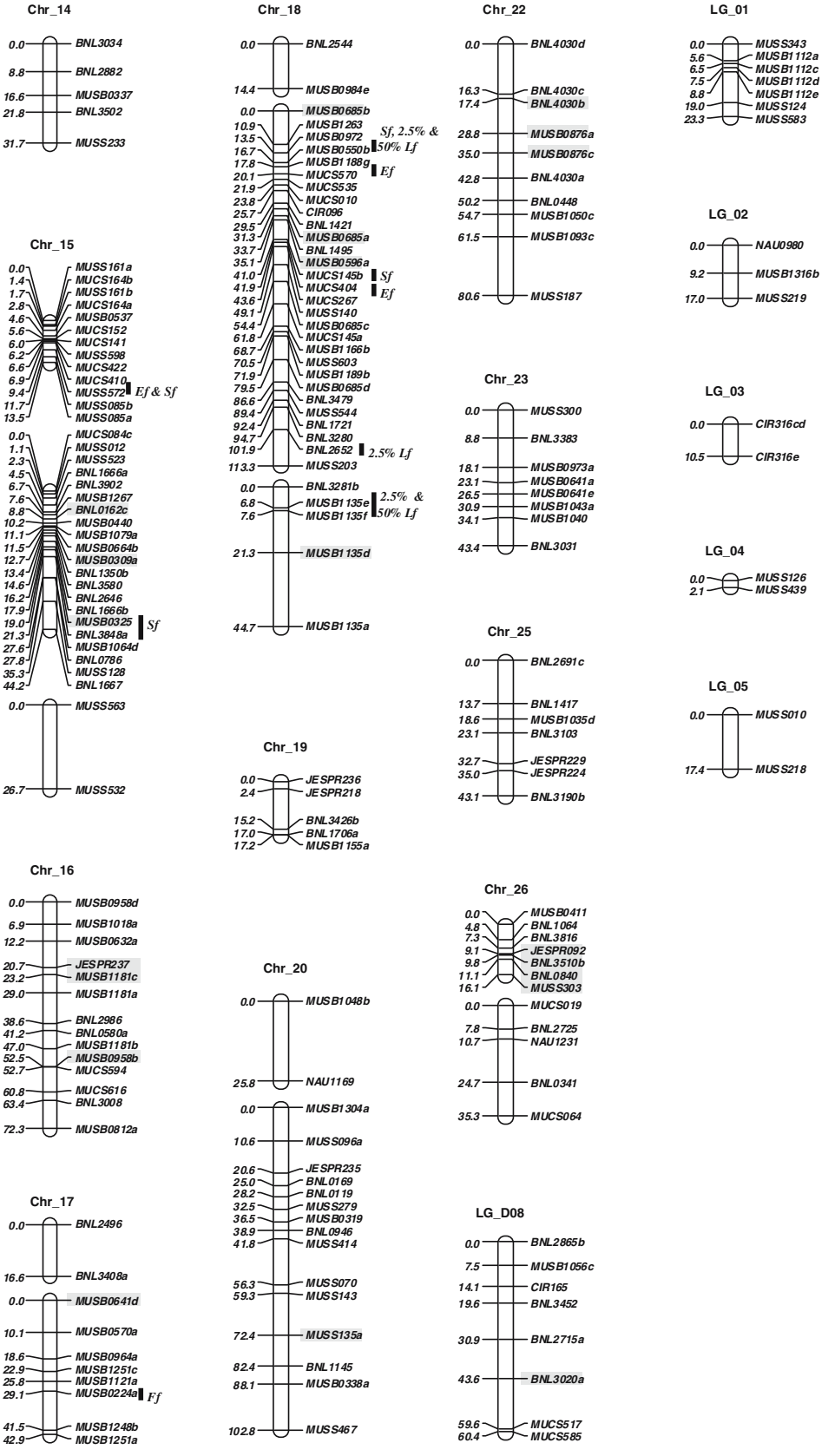

Fig. 2 D subgenome coverage from different chromosomes and unassigned linkage groups; the cotton genetic linkage map was constructed from 183 recombinant inbred lines (RILs) from the interspecific cross *Gh cv.* TM1 × *Gb acc.* Pima 3-79. from a total of 433 loci and 46 linkage groups, including 169 new BAC-MUSB loci and framework SSR markers (BNL, JESPR, CIR and NAU) to anchor the map. The markers with significant segregation distortion (Chi-square > 10) are indicated by *gray blocks*. Map intervals for putative QTL for five fiber quality-related traits, fiber elongation (*Ef*), fiber bundle strength (*Sf*), fiber length at 2.5% (*2.5% Lf*), fiber length at 50% (*50% Lf*), and fiber fineness (*Ff*), were presented by *solid vertical bars*

previous studies (Nguyen et al. 2004; Rong et al. 2004; Song et al. 2005) are in this linkage group, and this provides evidence that A02 corresponds to all or part of chromosome 8. Another linkage group commonly reported as A01 is associated with markers that are located on chromosome 18 by hypoaneuploid deficiency analysis and prior mapping studies. A01 is an A subgenome linkage group that is not yet assigned to a chromosome and chromosome 18 is a D subgenome chromosome that traditionally shared molecular marker alleles of duplicated loci with A01 (Lacape et al. 2003; Han et al. 2004; Nguyen et al. 2004; Rong et al. 2004; Lin et al. 2005; Song et al. 2005). Therefore, it is not surprising that some marker alleles of A01 became linked with chromosome 18 in this map. Chromosome 11 includes some markers that were originally assigned to linkage group A03 (BNL1034 and BNL1151; Han et al. 2004; Nguyen et al. 2004; Rong et al. 2004). A second linkage group also contains some markers assigned to A03 in these same studies, but assigning these markers in our map as part of chromosome 11 has been difficult as the markers failed to generate congruent results with the hypoaneuploid deficiency analysis. Further research is needed with additional markers and hypoaneuploid assays to further confirm these assignments.

Gene function categorization

Twenty-five mapped MUSB markers (presented in Table S2) have a match in protein homology in Genbank (BLASTX). Twenty-four showed matches to a plant sequence and six of these are associated with transposable or retroelements. Tomkins et al. (2001) reported frequent homology of BAC-end sequences to those sequences in the database identified as retroelements. The rest of the sequences appear to be a random set of protein homologies as might be expected when sequencing the genome at random. A bias of microsatellites with retroelements or the inherent frequencies of retroelements in genomic DNA are possible explanations of the results. A large proportion of the BAC-end sequences are homologous to unknown proteins or sequences that are yet to be characterized for protein synthesis or other functions (Tomkins et al. 2001). In addition this genetic map contains 86 MUSS and MUCS EST-derived markers that showed sequence homology to annotated genes with known function (Park et al. 2005) and 22 NAU markers (Han et al. 2004) which were originally designed from homologous sequences that were highly expressed, up- and down-regulated, in 24-day post-anthesis (dpa) cotton fibers relative to 10 dpa fiber (Arpat et al. 2004). Potentially, these sources of markers can help determine the contribution of candidate (or functionally characterized) sequences and unknown sequences in QTL analyses of fiber traits.

Fiber quality QTL analysis

Following the expansion of the map of Park et al. (2005) with MUSB and additional BNL markers, we sought to verify their previously detected QTLs because additional coverage of the cotton genetic map may confirm QTLs, or discover new ones, by added marker-trait variation. The phenotypic values for fiber elongation (*Ef*), bundle strength (*Sf*), length at 2.5% *Lf* and 50% *Lf* and fineness (*Ff*) were analyzed using MapQTL 4.0 with interval, MQM and restricted MQM mapping. The significant minimum LOD level for declaring a QTL was established at 2.9 for all five traits, determined by 1,000 permutation tests at $P < 0.05$. Table 4 summarizes the markers detected as QTLs with interval mapping, MQM and restricted MQM mapping, or both procedures along with their LOD scores and percent explanation of the trait variance. No QTLs were previously detected at the minimum LOD threshold for *Ef*, but our analysis showed a higher LOD score for MUSS572 (Park et al. 2005) and the appearance of another marker BNL3408, on chromosome 3. A QTL for *Sf* was detected on chromosome 18 for MUSB0685b with the minimum threshold cutoff. Two QTLs were detected on chromosome 18 for 2.5%*Lf* with MUSB0685b and BNL2652. Three QTLs for 50%*Lf* were marked with MUCS620 and MUSB0958e (chromosome 2), and MUSB0685b (chromosome 18). Two QTLs for *Ff* were marked with BNL3408 (chromosome 3) and MUSS026 (chromosome 12).

Discussion

Newly developed BAC-end MUSB markers improved genome coverage in the consensus genetic map and, for the first time by hypoaneuploid deficiency analysis, assigned several loci onto cotton chromosomes. MUSB-derived markers will also, for the first time, aid in the reconciliation of the genetic and physical maps, and fine genetic and QTL mapping of important loci on chromosomes that may affect variation in fiber quality traits, and pest resistance. Original BAC clones may provide critical targets of DNA regions for the discovery of gene sequences involved in biological processes such as fiber development and pest resistance in cotton. We designed 1,316 MUSB primer pairs and placed 169 MUSB marker loci and added 849 cM to the genetic distance into the map using RILs from the interspecific cross of TM-1 × 3-79 (Park et al. 2005). During the course of the SSR primer pair designs, we observed that the composition of the SSR motifs within our MUSB markers differed in type and frequency. There was a preponderance of AT, TA and AAG for di and trinucleotide motifs, similar to cotton ESTs (Park et al. 2005) and *Arabidopsis* (Cardle et al. 2000). However, other plant species such as wheat, rice and barley (La Rota et al. 2005) differ in their lower frequency of these motifs and a higher frequency of

Table 4 Quantitative trait loci (QTL) detected by MapQTL 4.0

Trait	QTL	Map location (subgenome)	Nearest marker	LOD	%expl
Fiber elongation	*Ef*	Chr.03 (A)	BNL3408b[a]	2.5	14.5
		Chr.15 (D)	MUSS572[c]	2.3	5.8
		Chr.18 (D)	MUCS404[a]	2.1	5.6
		Chr.18 (D)	MUCS570[a]	2.1	5.7
Fiber bundle strength	*Sf*	Chr.03 (A)	MUSB0073d[a]	2.1	5.7
		Chr.15 (D)	MUSB0325[c]	2.5	5.5
		Chr.15 (D)	MUSB1064c[c]	2.4	5.9
		Chr.15 (D)	MUSS572[a]	2.0	5.4
		Chr.18 (D)	MUCS145b[c]	2.6	6.3
		Chr.18 (D)	MUSB0685b[c]	2.9	6.5
Fiber length at 2.5%	2.5% *Lf*	Chr.02 (A)	MUSS294b	2.0	5.7
		Chr.09 (A)	MUSS022[c]	2.3	8.0
		Chr.18 (D)	MUSB1135h[b]	2.2	7.2
		Chr.18 (D)	MUSB0685b[c]	3.0	8.0
		Chr.18 (D)	BNL2652[c]	2.9	9.0
Fiber length at 50%	50% *Lf*	Chr.02 (A)	MUCS620[c]	3.8	9.4
		Chr.02 (A)	MUSB0958e[b]	2.9	7.3
		Chr.10 (A)	BNL256[b]	2.3	5.9
		Chr.18 (D)	MUSB1135d[b]	2.2	7.6
		Chr.18 (D)	MUSB0685b[c]	3.1	7.8
Fiber fineness	*Ff*	Chr.03 (A)	BNL226b[b]	2.1	7.9
		Chr.03 (A)	BNL3408b[c]	3.8	17.2
		Chr.05 (A)	MUCS108[b]	2.2	7.3
		Chr.12 (A)	MUSS026[c]	3.6	11.1
		Chr.17 (D)	MUSB0224a[b]	2.1	6.8

MapQTL mapping output were used to represent the values for LOD and phenotypic variance explained (%expl). Markers with LOD > 2.0 are included because some markers appear in more than one trait
[a]QTL detected by interval mapping
[b]QTLs detected by MQM mapping and restricted MQM mapping
[c]QTLs detected by interval mapping, MQM and restricted MQM mapping

CGC and CCG motifs. The bases C and G rarely co-occur in our SSR motifs and this may be a distinguishing feature between monocotyledonous and dicotyledonous plants. Park et al. (2005) reported that trimeric type of SSR motif was the most frequent in the EST SSR markers whereas we found that in BAC-ends tetrameric SSRs were more prevalent suggesting that genomic DNA, in general, does not select for the trimeric SSRs whereas some EST sequences may only tolerate variation in the trimeric SSR units (Metzgar et al. 2000).

MUSB markers not only expanded the length of the map, but also improved genome coverage by filling the gaps between loci as well. Park et al. (2005) had an average distance of 6.7 and 6.5 cM between loci in the At and Dt subgenomes but in this map (Figs. 1 and 2) they decreased to 4.9 and 5.0. This is a significant new source of markers because other maps with more markers still have a higher average distance between loci, e.g., Lin et al. 2005 (566 loci, 9.1 cM distance) and Song et al. 2005 (624 loci, 9.0 cM). However, some markers appear to be clustered in centromeric and telomeric regions and still leave gaps in the genetic maps, suggesting that the genomic map is currently composed of areas with high (hot spot) and low (cold spot) meiotic recombination like many other eukaryotic organisms (Gill et al. 1996; Petes 2001; Ulloa et al. 2005). Verification of the physical density of map regions with different recombination rates (genetic distances) will require knowledge of the underlying genomic sequences.

Apparent paralogous loci between chromosome 3 and 14, and 5 and 20 have been observed in other studies (Reinisch et al 1994; Rong et al. 2004). Sixteen BNL markers were assigned to just one locus that agreed with the published cotton SSR maps, but some or all of these maps show additional homoeologous or paralogous loci for these same markers. Marker duplication across or within chromosomes may reveal clues to chromosomal and whole genome evolution. Differences in marker location can first be resolved with the use of aneuploid stocks since they provide cytogenetic evidence of chromosomal location and help validate the marker location (Figs. 1 and 2). Continued use of BAC-end derived markers and hypoaneuploid deficiency analysis will be instrumental to resolving marker order, distances, and the extent of sequence duplication in the cotton genomes. Cotton is also more closely related to *Arabidopsis* than most other crop plants, and this will enhance comparative mapping and genomic research with this complete genome (Rong et al. 2005). The ordering of most markers in the genetic map presented herein are in agreement with Park et al. (2005) and with the exception of ten markers the placement of the BNL, CIR and JESPR SSRs are in general agreement with other maps (Lacape et al. 2003; Han et al. 2004; Nguyen et al. 2004; Rong et al. 2004; Lin et al. 2005; Song et al. 2005). BNL852, BNL1045 and BNL3580 are on chromosome 5, 12 and 15, respectively, in this map but in other maps they lie on linkage group D08, and on the homoeologous chromosomes 26 and 1, respectively (Nguyen et al. 2004; Song et al. 2005). Han et al. (2004) placed BNL1045 on chromosome 3, whereas Rong et al. (2004) mapped it to chromosome 22. JESPR218, JESPR236 and JESPR235 are on chromosomes 19, 19, and 20 respectively, in our map, but the first two are assigned to linkage group D08 and JESPR235 to D08 and chromosome 22 in Song et al. (2005). CIR203 and BNL1064

are on linkage group A03 and chromosome 26, respectively, on this map, but CIR203 is on chromosome 6 in Nguyen et al. (2004) and BNL1064 is a consensus for chromosome 6 (Lacape et al. 2003; Han et al. 2004; Nguyen et al. 2004; Rong et al. 2004; Lin et al. 2005; Song et al. 2005). We mapped NAU0998a and NAU0922 to chromosome 3 and 5 but Han et al. (2004) placed these on chromosomes 14 and 20.

Distorted segregation of loci was detected for SSR markers, and these were skewed mainly towards the maternal parent (TM-1). This is commonly observed in wide crosses in cotton (Lacape et al. 2003; Han et al. 2004) and other plants (Tanksley et al. 1992) but the congruence of skewed markers in our map with other genetic maps (Han et al. 2004; Nguyen et al. 2004; Song et al. 2005) and the chromosomal assignments with hypoaneuploid deficiency analyses justifies their inclusion into this genetic map. Hypoaneuploid deficiency analyses and genetic linkage map construction suggested the localization of several SSR markers to chromosomes 8, 11, 19 and 22 which have not been previously reported (Figs. 1 and 2). These new findings provided new insights to the genome tetraploid composition in this genetic map.

The diploid A and D genomes, which differ in physical size by a factor of about 2, are recombinationally equivalent (Brubaker et al. 1999), as are the At and Dt tetraploid subgenomes (Brubaker et al. 1999; Rong et al. 2004). However, the genetic distances of the At and Dt subgenomes are about 50–93% longer than their extant diploid models (Reinisch et al. 1994; Brubaker et al. 1999; Rong et al. 2004). Markers specific to tetraploid cotton such as the MUSB markers will be needed to cover this extra distance. Based on genome-specific chromosomes identified in *G. hirsutum* tetraploid (A and D), the A subgenome in our genetic map covers 56.9% of the genome while the D subgenome covers 39.7%. Also more MUSB markers were linked to the A subgenome (105) than the D subgenome (57) (Figs. 1 and 2), which suggests that the A subgenome is more polymorphic and recombinationally active than originally thought (Zhao et al. 1998). Frequent association of repetitive elements in BAC-end sequences in general (Tomkins et al. 2001) and MUSB markers (supplementary Table S2) suggest that larger genome sizes may result from retroelement activity (Zhao et al. 1998; Zaki and Ghany 2004), which has been supported by research in the rice (Ma et al. 2004) and human (Boissinot et al. 2004) genomes. The MUSB markers appear to be targeting segments of both genomes in proportion to their physical size, thus helping to reconcile genetic map distances with physical distances. Continued research with BAC-end markers and alignment of BAC contigs will also enable research into the causes of different genome and subgenome sizes among related species.

This map was useful in detecting fiber quality QTLs similar to Park et al. (2005) and Kohel et al. (2001). More precise chromosomal locations were detected for QTLs in this study than Park et al. (2005) and demonstrated the utility of adding new MUSB SSR markers. Table 4 includes markers with LOD > 2.0 because some markers appear in more than one trait, e.g., BNL3408 (chromosome 3), which explains 17.2 and 14.5% of the variation in *Ff* and *Ef*, respectively. The same markers described in Park et al. (2005) were detected here, but the LOD scores and percent variance explained were increased in our map for 50% *Lf* (MUCS620) and *Ff* (BNL226b and BNL3408b). Marker MUCS620 is homologous to an endo-β-1,4-glucanase (*E* value = 0.0), an enzyme that plays an important role in cell extension during rapid polar elongation of developing fibers (Arpat et al. 2004), but the discovery of additional QTLs in our map suggests that DNA marker sequences on other chromosomes also affect fiber length. In general, markers that showed the highest LOD scores were detected in multiple traits (e.g., MUSB0685b on chromosome 18 explained 6.5, 9.0 and 7.8% variation in *Sf*, 2.5% *Lf* and 50% *Lf*, respectively). Kohel et al. (2001) also reported fiber QTLs on chromosome 3 and 18 which parallel results of other studies that implicate these chromosomes and others in controlling fiber quality (Lacape et al. 2003; Ulloa et al. 2005). Research reported by Morgante et al. (2002) and La Rota et al. (2005) suggested that genomic microsatellites were preferentially associated with gene-rich regions. These results justify the use of BAC-end MUSB markers along with EST-SSR markers to fine map QTLs. Saturation of the cotton genetic map with BAC-end markers will further improve the targeting of common QTLs for discovery of genes and their biological roles in complex traits and to complement research in cotton biology (Wang et al. 2004; Wilkins and Arpat 2005).

The assignment of chromosome location to markers still associated with subgenomes (e.g. linkage A03) has been difficult. Hypoaneuploid deficiency analysis in this study has enabled the assignment of some SSR markers to chromosome 11 that were originally linked to A03. Important disease resistance traits have recently been associated with SSR markers in linkage group A03 and chromosome 11 (Bolek et al. 2005; Wang et al. 2005). Continued use of genomic tools such as hypoaneuploid deficiency analysis, use of genomic markers such as MUSB BAC-derived SSRs, and a genetic framework, map, will enable the discovery of gene sequences involved in disease resistances to several common soil pathogens and facilitating the fine mapping of these genes.

Knowledge of the actual relationships between physical and genetic distances, and references to whole genomic DNA is lacking in cotton. Kim et al. (2005) used fluorescence in situ hybridization (FISH) of BAC contigs within the framework of the *Sorghum* genetic map and found that the new estimates of physical distances was 1-fold less than the estimates derived from genetic distances. The concurrent development of sequence-tagged connectors and BAC fingerprints (Tomkins et al. 2001; Yu et al. 2005) in the course of BAC-end sequencing and the placement of BAC-end MUSB markers into the genetic map will contribute to the

development of a consensus map and enable the physical alignment of the cotton genome. This will improve the assessment of positional cloning and fine mapping for genes of interest and enhance investigations into their genetic sequences and biological roles in the cotton plant. An additional benefit is that the MUSB markers are helpful in the development of a consensus PCR-based map and provide the cotton community with marker-assisted tools for cotton breeders to discover and select for genes to improve agronomic, fiber and pest resistance traits.

Acknowledgements We would like to gratefully acknowledge Cotton Inc. for its support on this project under a Cotton Incorporated Fellowship (Cary, NC) for James Frelichowski. The sequences of these markers are made public through the Cotton Microsatellite Database (CMD, http://www.mainlab.clemson.edu/cmd/projects/) as part of a community-wide initiative to increase coverage of the cotton genome. Names are necessary to report factually in available data; however, the USDA neither guarantees nor warrants the standard of products or service, and the use of the name by the USDA implies no approval of the products or service to the exclusion of others that may also be suitable. This research was supported in part, by Cotton Incorporated Project Number 05-704 (to MU) under the C.I. Fellowship program.

References

Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Anderson CG (1999) Cotton marketing. In: Smith CW, Cothren JT (eds) Cotton; origin, history, technology, and production. Wiley, New York, pp 659–679

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage map. Genome 36:181–186

Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D, Main D, Wood T, Leslie A, Wing RA, Wilkins TA (2004) Functional genomics of cell elongation in developing cotton fibers. Plant Mol Biol 54:911–929

Boissinot S, Entezam A, Young L, Munson PJ, Furano AV (2004) The insertional history of an active family of L1 retrotransposons in humans. Genome Res 14:1221–1231

Bolek Y, Kamal M El-Zik, Pepper AE, Bell AA, Magill CW, Thaxton PM, Reddy OUK (2005) Mapping of verticillium wilt resistance genes in cotton. Plant Sci 168:1581–1590

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet 32:314–331

Brubaker CL, Paterson AH, Wendell JF (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42:184–203

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative traits mapping. Genetics 138:963–971

De Kock MJD, Brandwagt BF, Bonnema G, de Wit PJGM, Lindhout P (2005) The tomato Orion locus comprises a unique class of *Hcr9* genes. Mol Breed 15:409–422

Fryxell PA (1992) A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). Rheedea 2:108–165

Gill KS, Bill BS, Endo TR, Taylor T (1996) Identification and high density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics 144:1883–1891

Gonzalez J, Nefedov M, Bosdet I, Casals F, Calvete O, Delprat A, Shin H, Readman C, Mathewson C, Wye N, Hoskins RA, Schein JE, deJong P, Ruiz A (2005) A BAC-based physical map of the *Drosophila buzzatii* genome. Genome Res 15:885–892

Gregory SR, Hernandez E, Savoy BR (1999) Cottonseed processing. In: Smith CW, Cothren JT (eds) Cotton; origin, history, technology, and production. Wiley, New York, pp 793–823

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Current Sci 70:45–53

Han Z, Guo W, Song X, Zhang T (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol Genet Genomics 272:308–327

Hof JV, Saha S (1997) Cotton fibers can undergo cell division. Am J Bot 84(9):1231–1235

Hong CP, Lee SJ, Park JY, Plaha P, Park YS, Lee YK, Choi JE, Kim KY, Lee JH, Lee J, Jin H, Choi SR, Lim YP (2004) Construction of a BAC library of Korean ginseng and initial analysis of BAC-end sequences. Mol Genet Genomics 271:709–716

Islam-Faradi MN, Childs KL, Klein PE, Hodnett G, Menz MA, Klein RR, Rooney WL, Mullet JE, Stelly DM (2002) A molecular cytogenetic map of sorghum chromosome 1. Fluorescence *in situ* hybridization analysis with mapped bacterial artificial chromosomes. Genetics 161:345–353

Jiang C, Wright RJ, El-Zik KM, Paterson AH (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). Proc Natl Acad Sci 95:4419–4424

Kalendar R (2005) FastPCR: a PCR primer design and repeat sequence searching software with additional tools for the manipulation and analysis of DNA and protein (http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm)

Kilian A, Chen J, Han F, Steffenson B, Kleinhofs A (1997) Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. Plant Mol Biol 35:187–195

Kim HJ, Triplett BA (2001) Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. Plant Phys 127:1361–1366

Kim J-S, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005) Molecular cytogenetic maps of Sorghum linkage groups 2 and 8. Genetics 169:955–965

Kohel RJ, Yu J, Park Y-H, Lazo GR (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica 121:163–172

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Lacape J-M, Nguyen T-B (2005) Mapping quantitative trait loci associated with leaf and stem pubescence in cotton. J Hered 96(4):441–444

Lacape J-M, Nguyen T-B, Thibivilliers S, Bojinov B, Courtois B, Cantrell RG (2003) A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* × *Gossypium barbadense* backcross population. Genome 46:612–626

La Rota M, Kantety RV, Yu J-K, Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellites markers in rice, wheat, and barley. BMC Genomics 6:23 (http://www.biomedcentral.com/1471-2164/6/23) DOI: 10.1186/1471-2164-6-23

Lin Z, He D, Zhang X, Nie Y, Guo X, Feng C, Stewart J McD (2005) Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD. Plant Breed 124:180–187

Liu S, Saha S, Stelly D, Burr B, Cantrell RG (2000) Chromosomal assignment of microsatellite loci in cotton. J Hered 91:326–332

Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14:860–869

Mao L, Wood TC, Yu Y, Budiman MA, Tomkins J, Woo S, Sasinowski M, Presting G, Frisch D, Goff S, Dean RA, Wing RA (2000) Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. Genome Res 10:982–990

Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). Theor Appl Genet 108:280–291

Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutation limits microsatellite expansion in coding DNA. Genome Res 10:72–80

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genet 30:194–200

Nievergelt CM, Smith DW, Kohlenberg JB, Schork NJ (2004) Large-scale integration of human genetic and physical maps. Genome Res 14:1199–1205

Nguyen TB, Giband M, Brottier P, Risterucci A-M, Lacape J-M (2004) Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. Theor Appl Genet 109:167–175

Palmer MB, Main D, Frelichowski JE, Tomkins JP, Ulloa M (2004) High-throughput development of new molecular markers for cotton. National Cotton Council Beltwide Cotton Conference p 1131

Park Y-H, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, El-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred (RIL) cotton population. Mol Genet Genomics 274:428–441

Paterson AH (2002) What has QTL mapping taught us about plant domestication. New Phytol 154:591–608

Percival AE, Wendel JF, Stewart JM (1999) Taxonomy and germplasm resources. In: Smith CW, Cothren JT (eds) Cotton; origin, history, technology, and production. Wiley, New York, pp 33–63

Petes TD (2001) Meiotic recombination hot spots and cold spots. Nat Rev Genet 2:360–369

Qureshi SN, Saha S, Kantety RV, Jenkins JN (2004) EST-SSR: a new class of genetic markers in cotton. J Cotton Sci 8:112–123

Reinisch A, Dong J-M, Brubaker C, Stelly D, Wendel J, Paterson AH (1994) A detailed RFLP map of cotton (*Gossypium hirsutum* × *G. barbadense*); chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding X, Garza JJ, Marler BS, Park C, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao X, Zhu L, Paterson AH (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389–417

Rong J, Bowers JE, Schulze SR, Waghmare VN, Rogers CJ, Pierce GJ, Zhang H, Estill JC, Paterson AH (2005) Comparative genomics of *Gossypium* and *Arabidopsis*: unraveling the consequences of both ancient and recent polyploidy. Genome Res 15:1198–1210

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols, in the series methods in molecular biology. Humana Press, pp 365–386. Code available at http://www.fokker.wi.mit.edu/primer3/

Saha S, Karaca M, Jenkins JN, Zipf AE, Reddy OUK, Kantety RV (2003) Simple sequence repeats as useful resources to study transcribed genes of cotton. Euphytica 130:355–364

Shen X, Guo W, Zhu X, Yuan Y, Yu JZ, Kohel RJ, Zhang T (2005) Molecular mapping of QTLs for qualities in three diverse lines in Upland cotton using SSR markers 15:169–181

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachirii Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci 89:8794–8797

Song X, Wang K, Guo W, Zhang J, Zhang T (2005) A comparison of genetic maps constructed from haploid and BC1 mapping populations from the same crossing between *Gossypium hirsutum* L. and *Gossypium barbadense* L. Genome 48:378–390

Stam P, Van Ooijen JW (1995) JoinMap™ version 2.0: software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen

Stelly DM (1993) Interfacing cytogenetics with the cotton genome mapping effort. In: Herber DJ, Richter DA (eds) Beltwide Cotton Conference, January 10–14, 1993, New Orleans, LA. Memphis, TN: National Cotton Council of America; pp 1545–1550

Taliercio EW, Ulloa M (2003) The DNA sequence of a *Gypsy* element from *Gossypium hirsutum* L. and characterization of *Gypsy* elements in three *Gossypium* species. DNA sequence V 14(4):319–325

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roeder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160

Tautz D, Schotterer C (1994) Simple sequences. Curr Opin Genet 4:832–837

Tomkins JP, Peterson DG, Yang TJ, Main D, Wilkins TA, Paterson AH, Wing RA (2001) Development of genomic resources for cotton (*Gossypium hirsutum* L.): BAC library construction, preliminary STC analysis, and identification of clones associated with fiber development. Mol Breed 8(3):255–261

Ulloa M, Meredith WR Jr (2000) Genetic linkage map and QTL analysis of agronomic and fiber quality traits in an intraspecific population. J Cotton Sci 4:161–170

Ulloa M, Meredith WR Jr, Shappley ZW, Kahler AL (2002) RFLP genetic linkage maps from $F_{2.3}$ populations and a joinmap of *Gossypium hirsutum* L. Theor Appl Genet 104:200–208

Ulloa M, Stewart J McD, Garcia CE, Godoy AS, Gaytan-MA, Acosta-NS (2006) Cotton genetic resources in the western states of Mexico; *In situ* conservation status and germplasm collection for *ex situ* preservation. Genet Resour Crop Evol (in press). DOI 10.1007/s10722-004-2988-0

Ulloa M, Saha S, Jenkins JN, Meredith WR, McCarty JC, Stelly MD (2005) Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) joinmap. J Hered 96:132–144

Van Ooijen JW, Maliepaard C (1996) MapQTL™ Version 3.0, software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Wang S, Wang J-W, Yu N, Li C-H, Luo B, Gou J-Y, Wang L-J, Chen X-Y (2004) Control of plant trichome development by a cotton fiber MYB gene. The Plant Cell 16:2323–2334

Wang C, Ulloa M, Roberts PA (2005) Identification and mapping of microsatellite markers linked to the root-knot nematode resistance gene *rkn1* in Acala NemX (*Gossypium hirsutum* L.). Theor Appl Genet (in press) DOI:10.1007/S00122-005-0183

Wilkins TA, Arpat AB (2005) The cotton fiber transcriptome. Physiologia Plantarum 124:295–300

Yu, J, Kohel RJ, Xu Z, Dong J, Zhang H, Stelly DM, Zhu Y, Covaleda L (2005) Physical mapping of fiber development genes in cotton [abstract]. Plant And Animal Genome XIII Conference p 225

Zaki EA, Ghany AAA (2004) Ty3/*gypsy* retro-transposons in Egyptian cotton (*G. barbadense*). J Cotton Sci 8:179–185

Zhao X, Si Y, Sason RE, Crane CF, Price HJ, Stelly DM, Wendell JF, Paterson AH (1998) Dispersed repetitive DNA has spread to new genomes since polyploidy formation in cotton. Genome Res 8:479–492